www.bsc.es

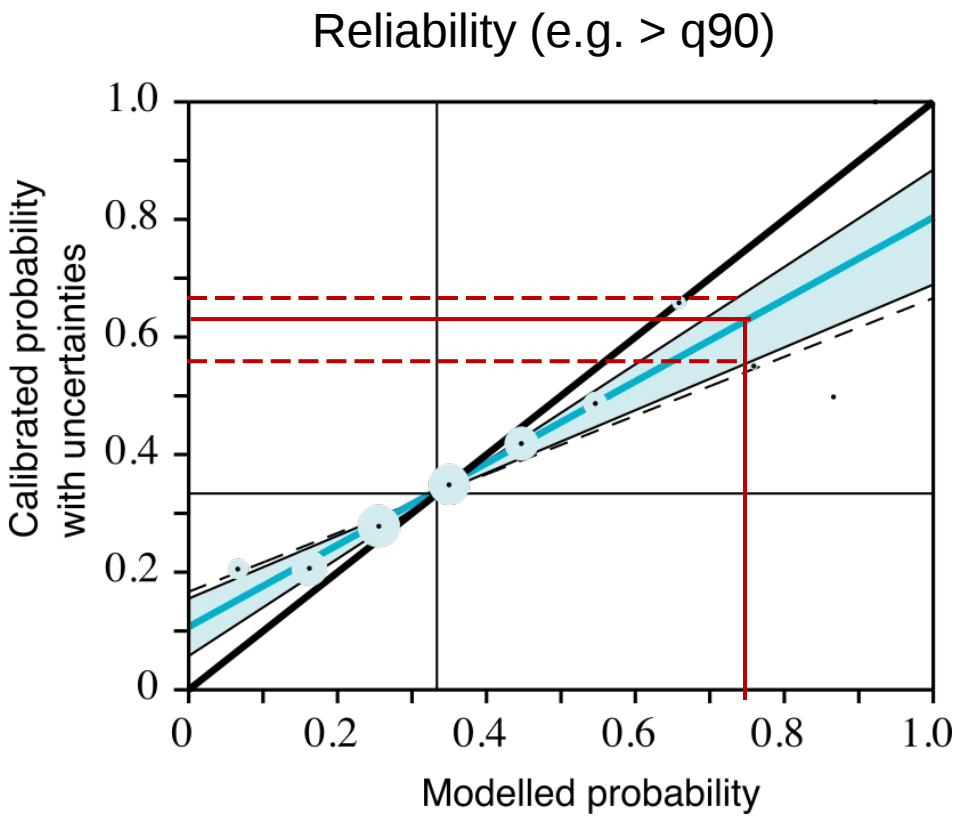
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Calibration of event probabilities

Omar Bellprat, Fraser Lott, Francisco Doblas-Reyes

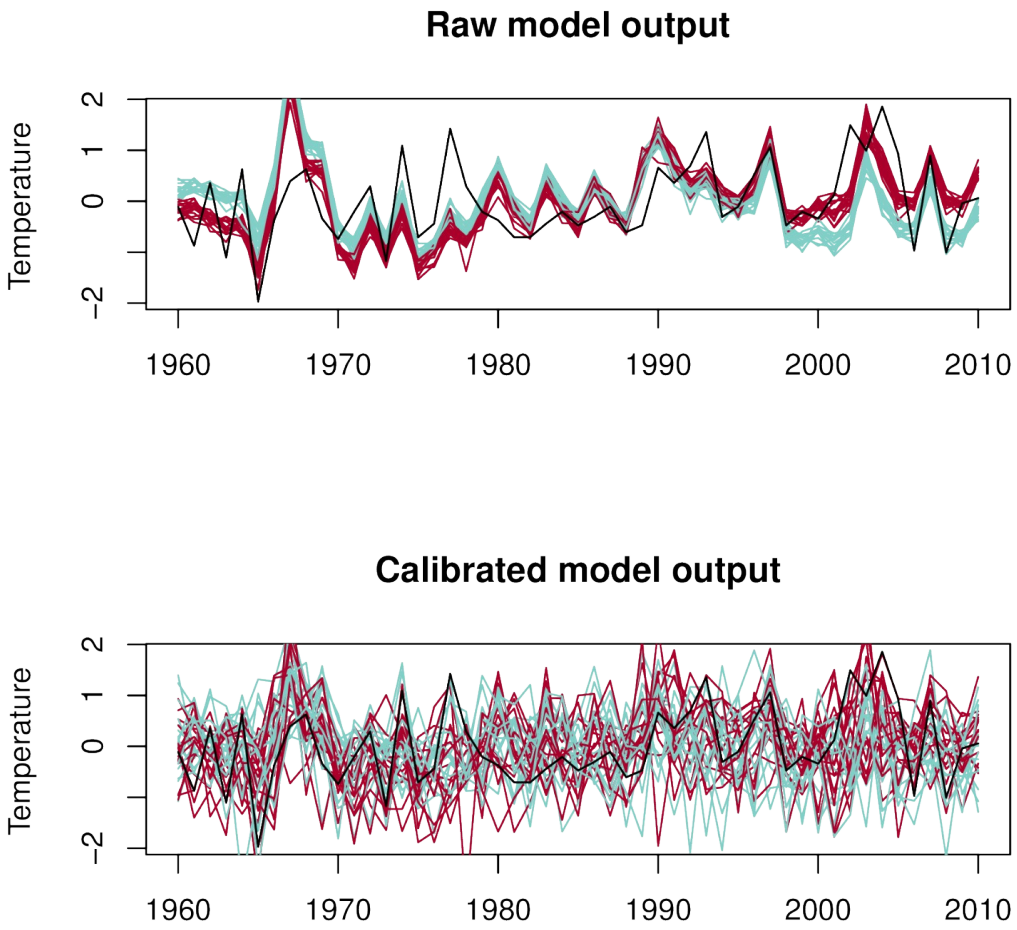
EUCLEIA GA, MetOffice, 30 November, 2016

Calibration: Reliability diagram



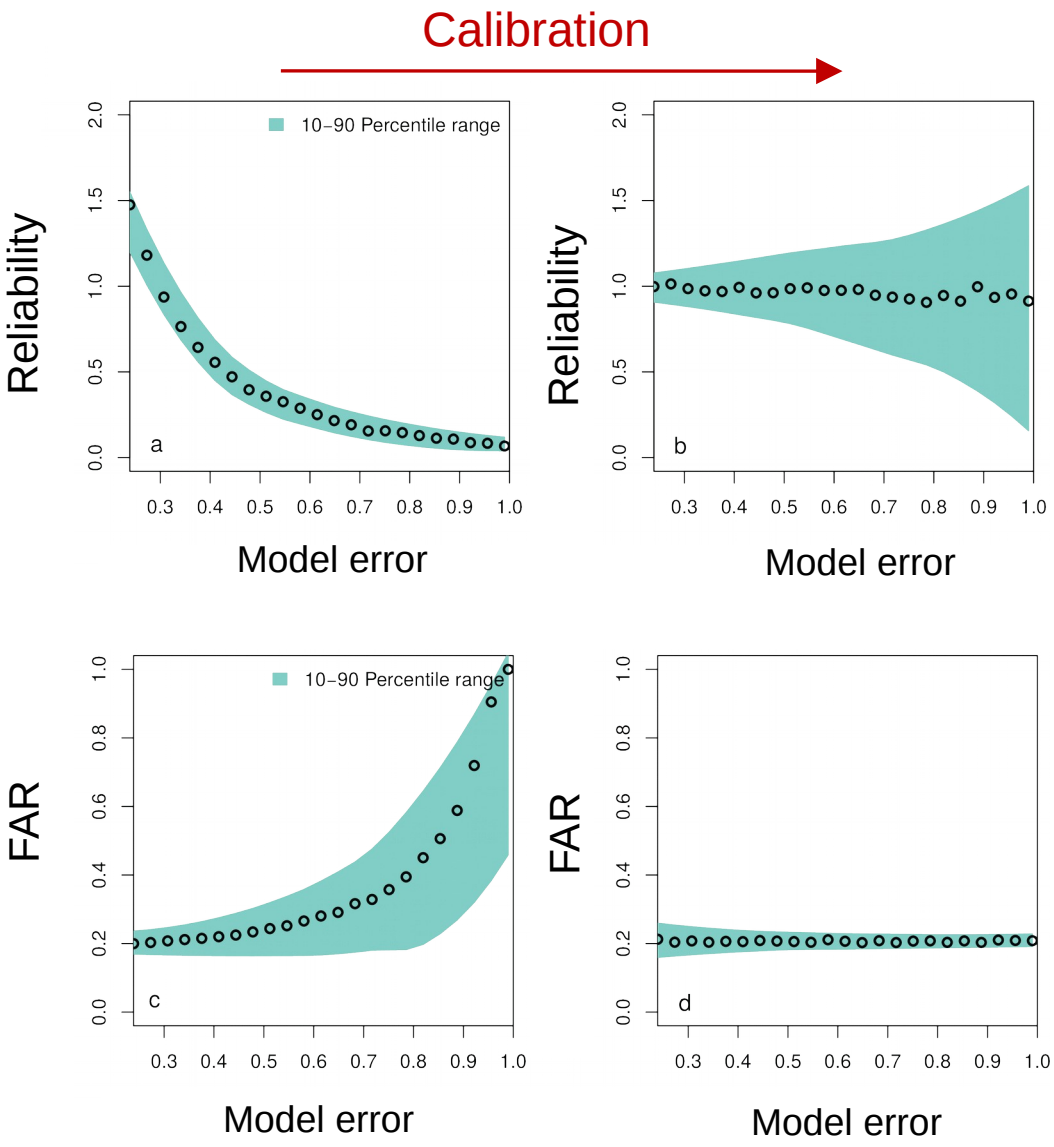
Palmer et al. (2008), Weisheimer and Palmer (2014)

Calibration: Ensemble inflation



Doblas-Reyes et al. (2005), Bellprat and Doblas-Reyes (2016)

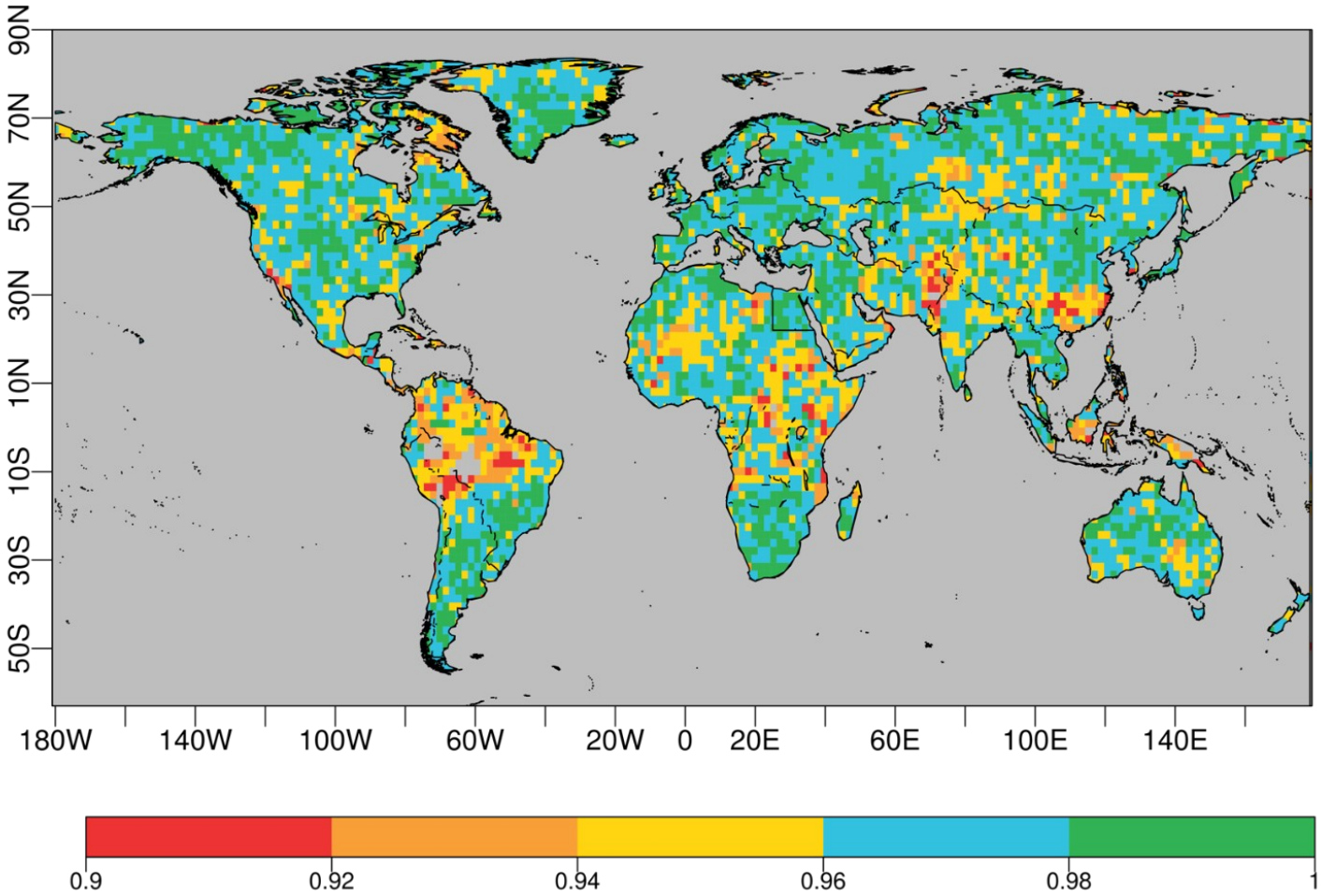
Calibration a statistical model



Bellprat and Doblas-Reyes (2016), WP 6.4 Deliverable (2016)

Reliability HadGEM3-A

Reliability of hot summers (> q90) 1961-2010

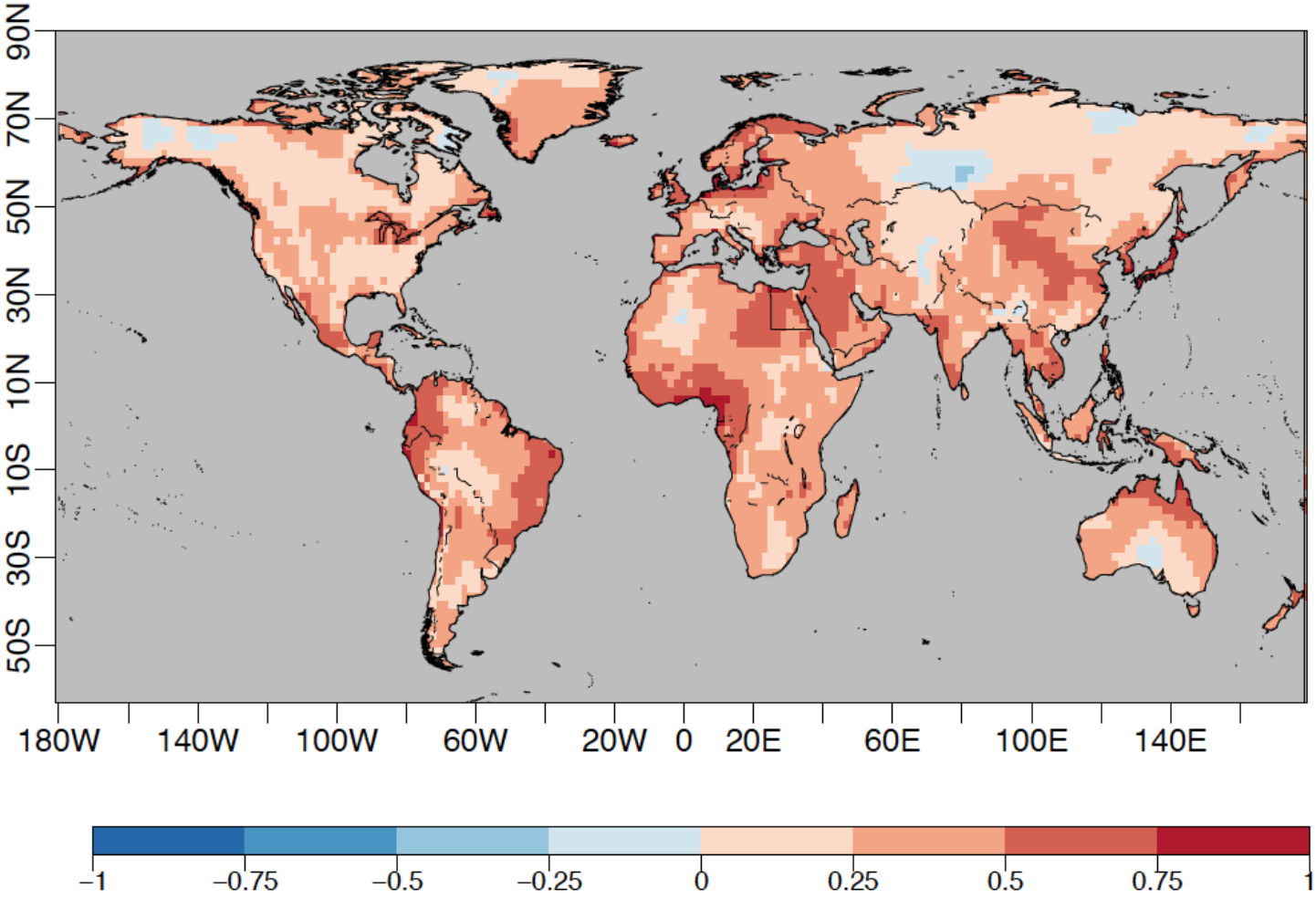


Reliability (Brier-component)

e.g van Oldenborgh et al. (2014)

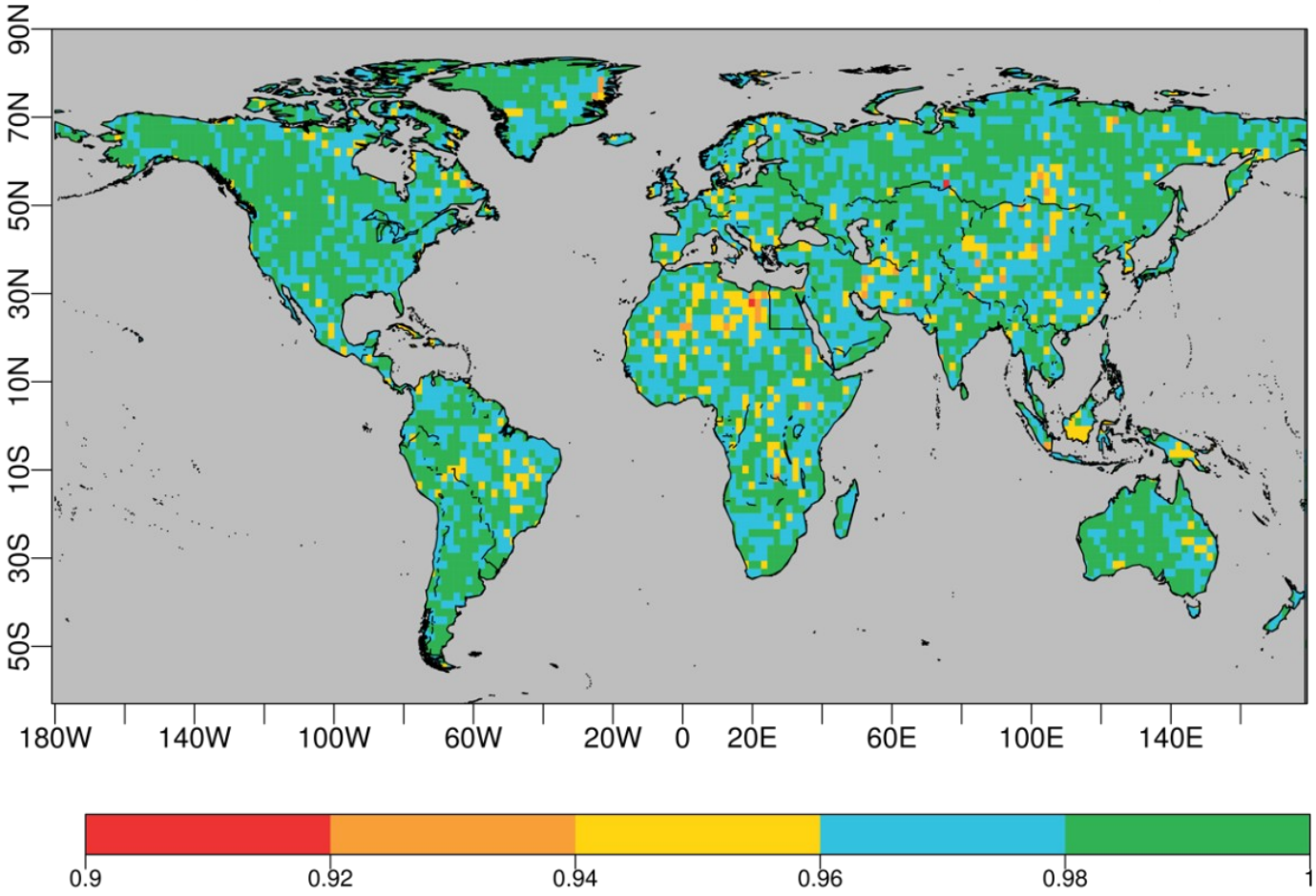
“Skill” HadGEM3-A

Detrended ensemble mean correlation mean JJA temperature



Calibrated HadGEM3

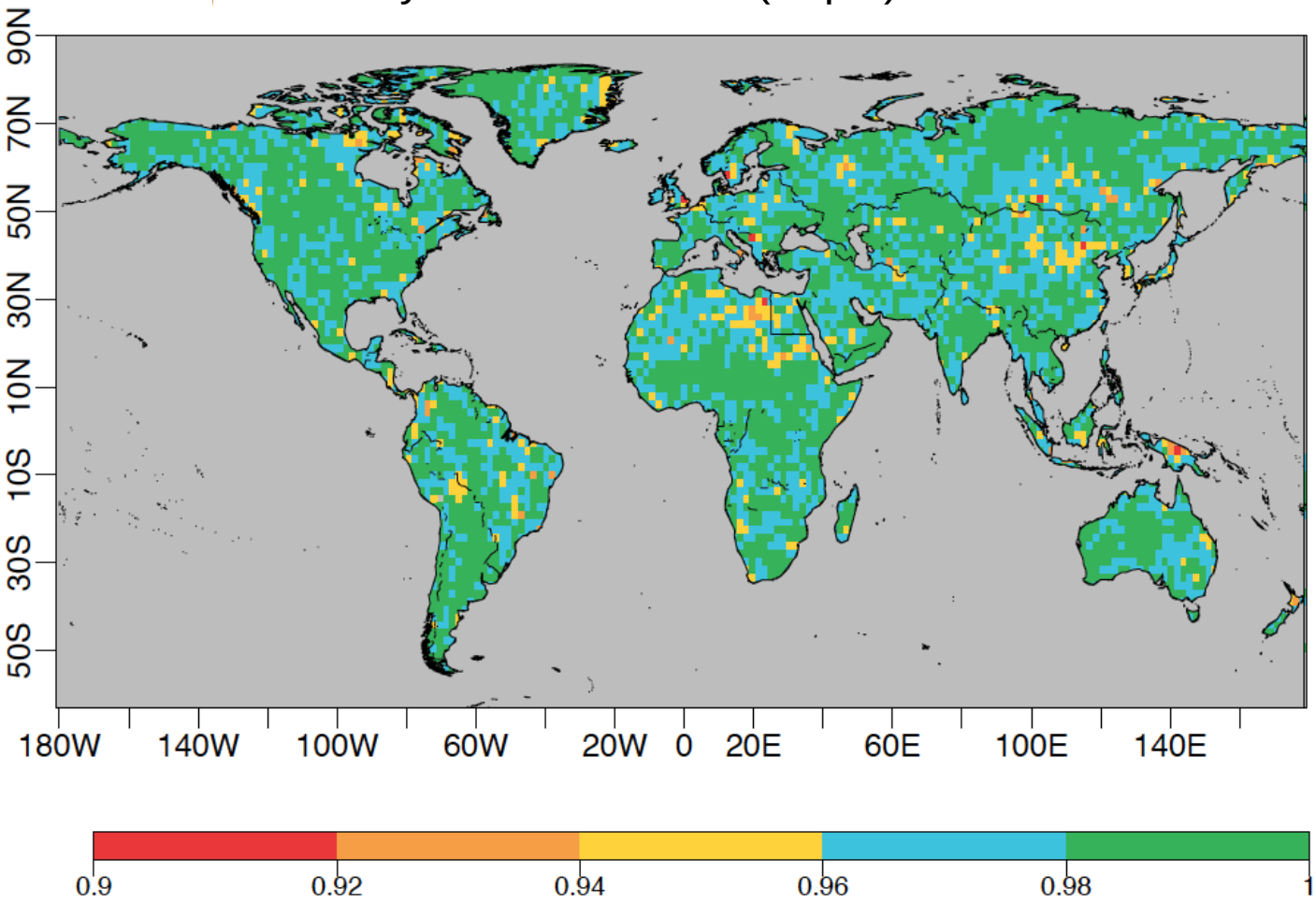
Reliability of hot summers (> q90) 1961-2010



Reliability (Brier-component)

C20C+ Multi-model AMIP

Reliability of hot summers (> q90) 1961-2010

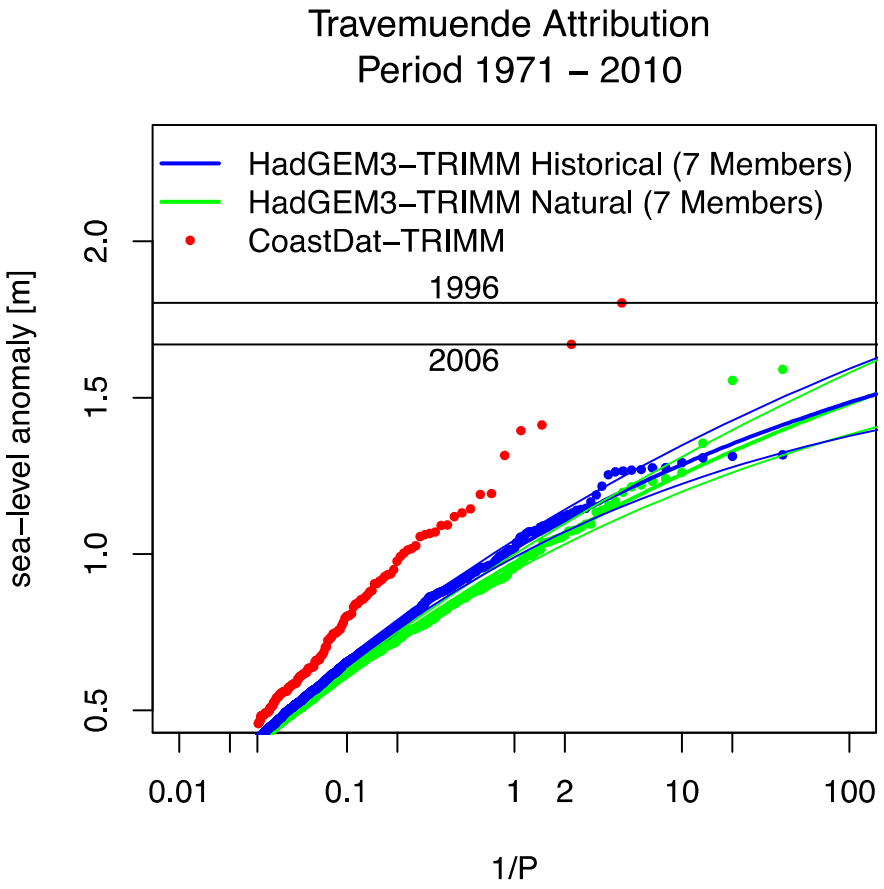


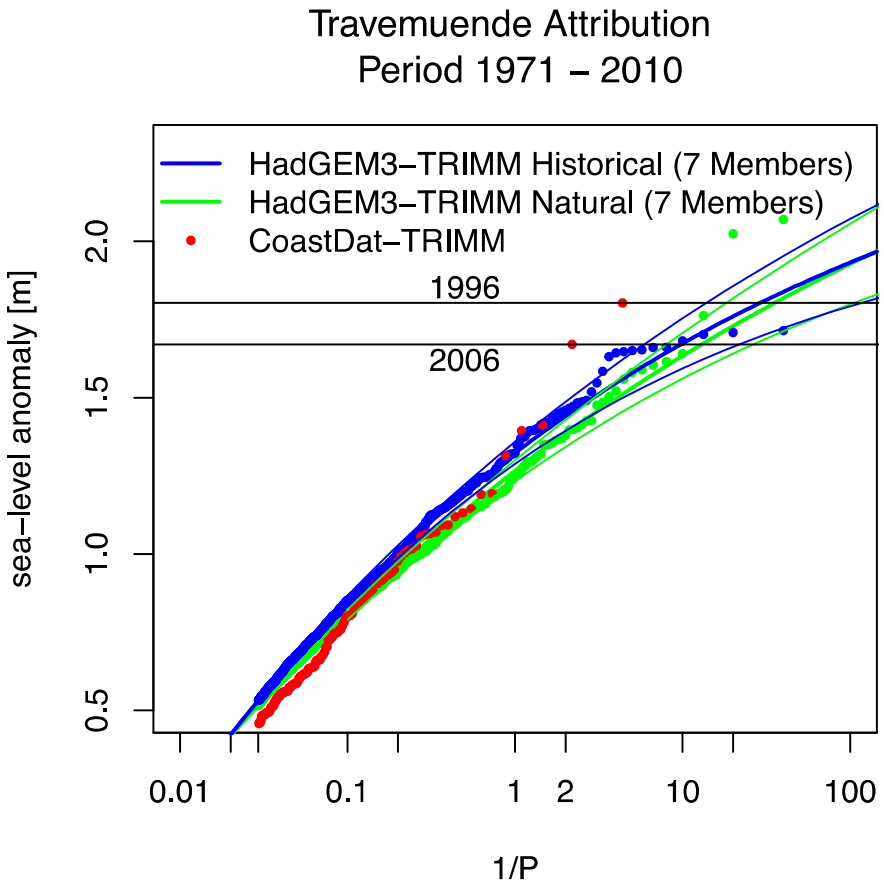
Reliability (Brier-component)

Stone et al.

Pooling years and members + bias-correction

Interannual variability and ensemble spread less an issue.



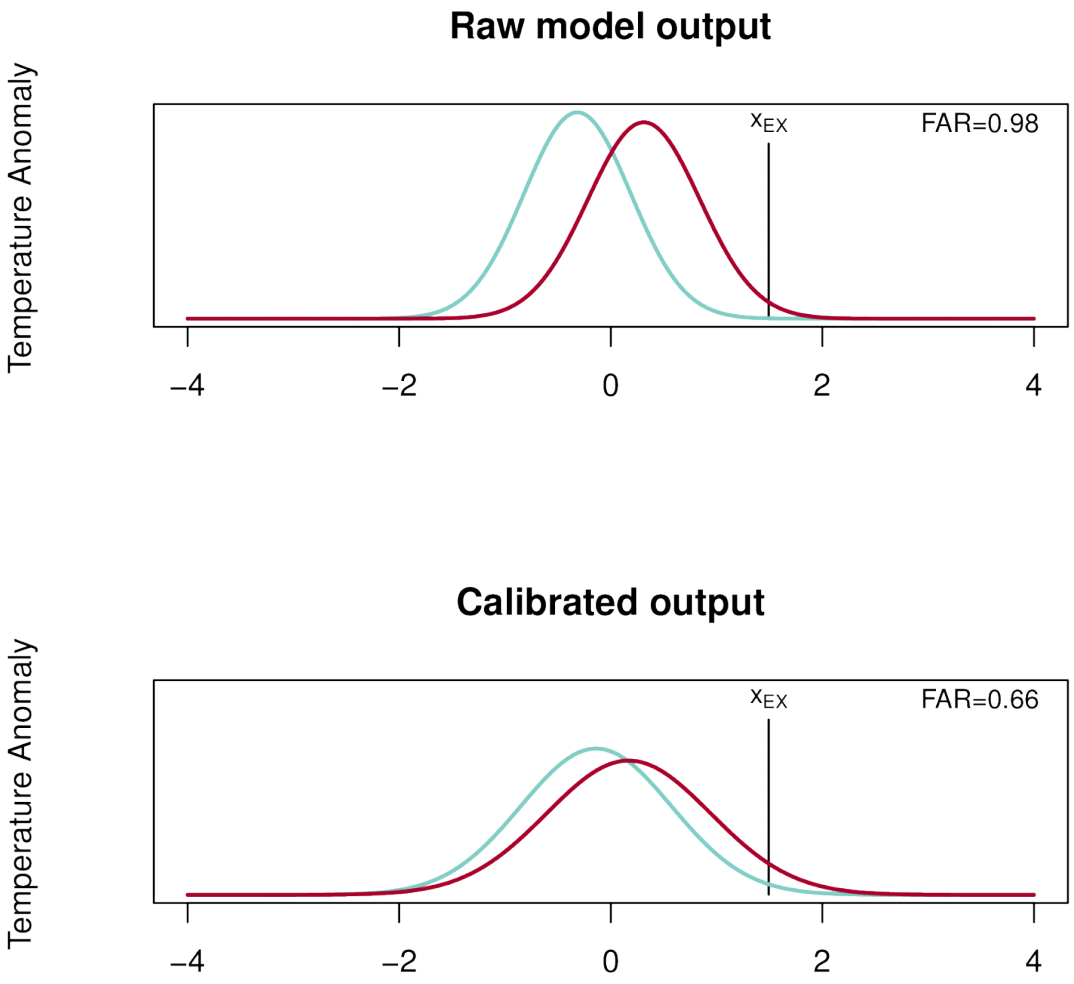


Klehmet et al., Storm Surge Case

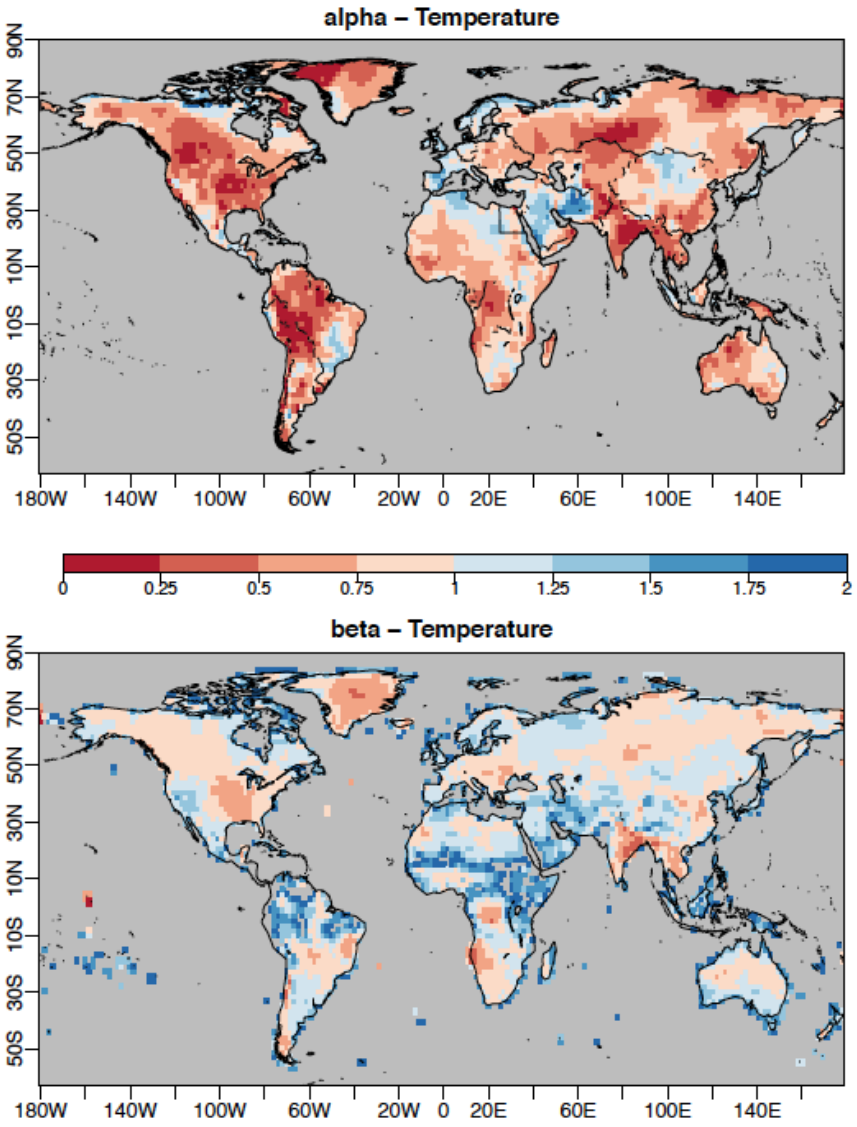
Summary

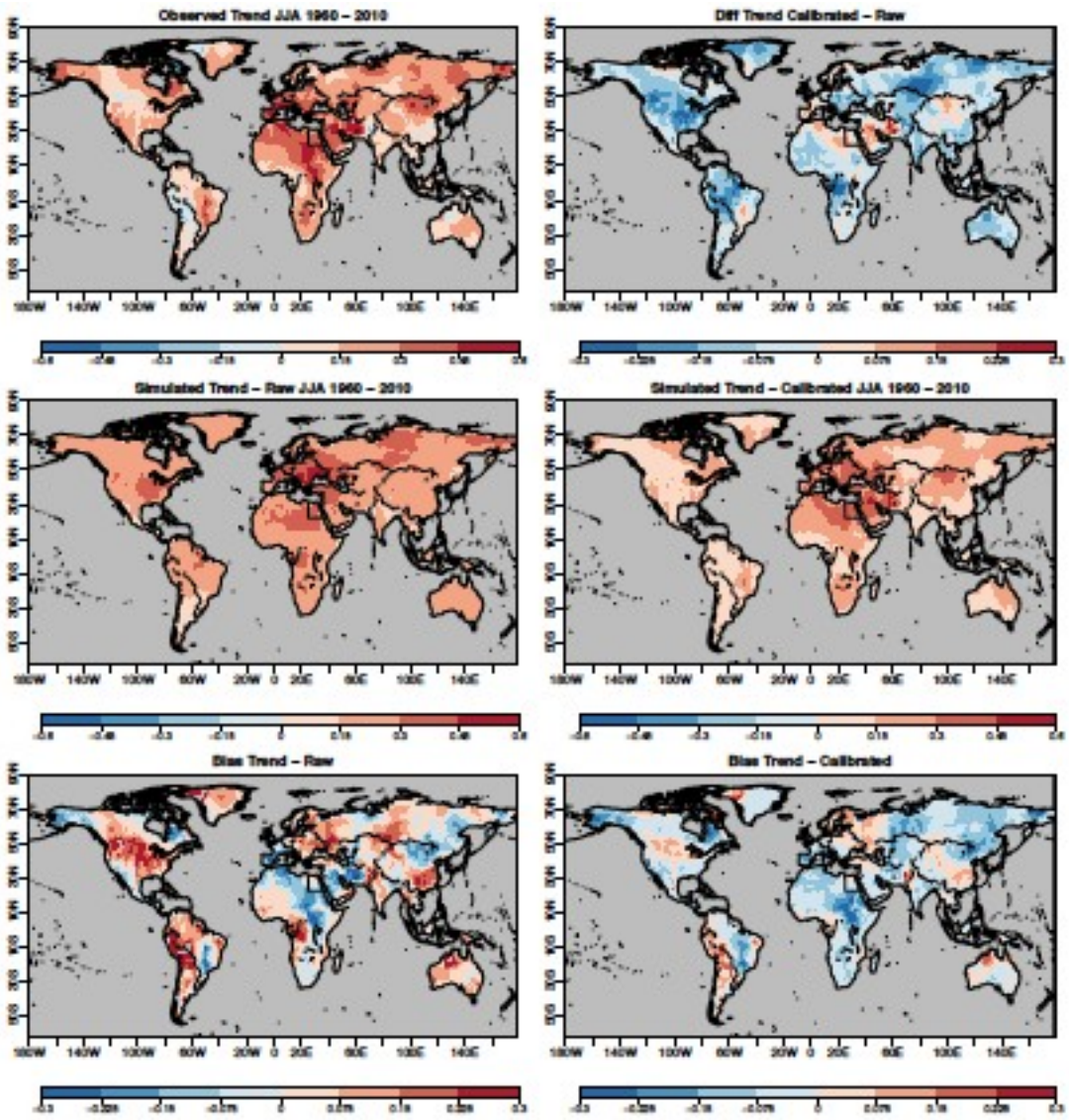
- Ensemble calibration improves accuracy of event attribution statements
- Multi-model ensembles and pooling of years and members in combination with bias-correction improves reliability.
- How much calibration + bias-correction should we trust?

Calibration reduces attributable risk

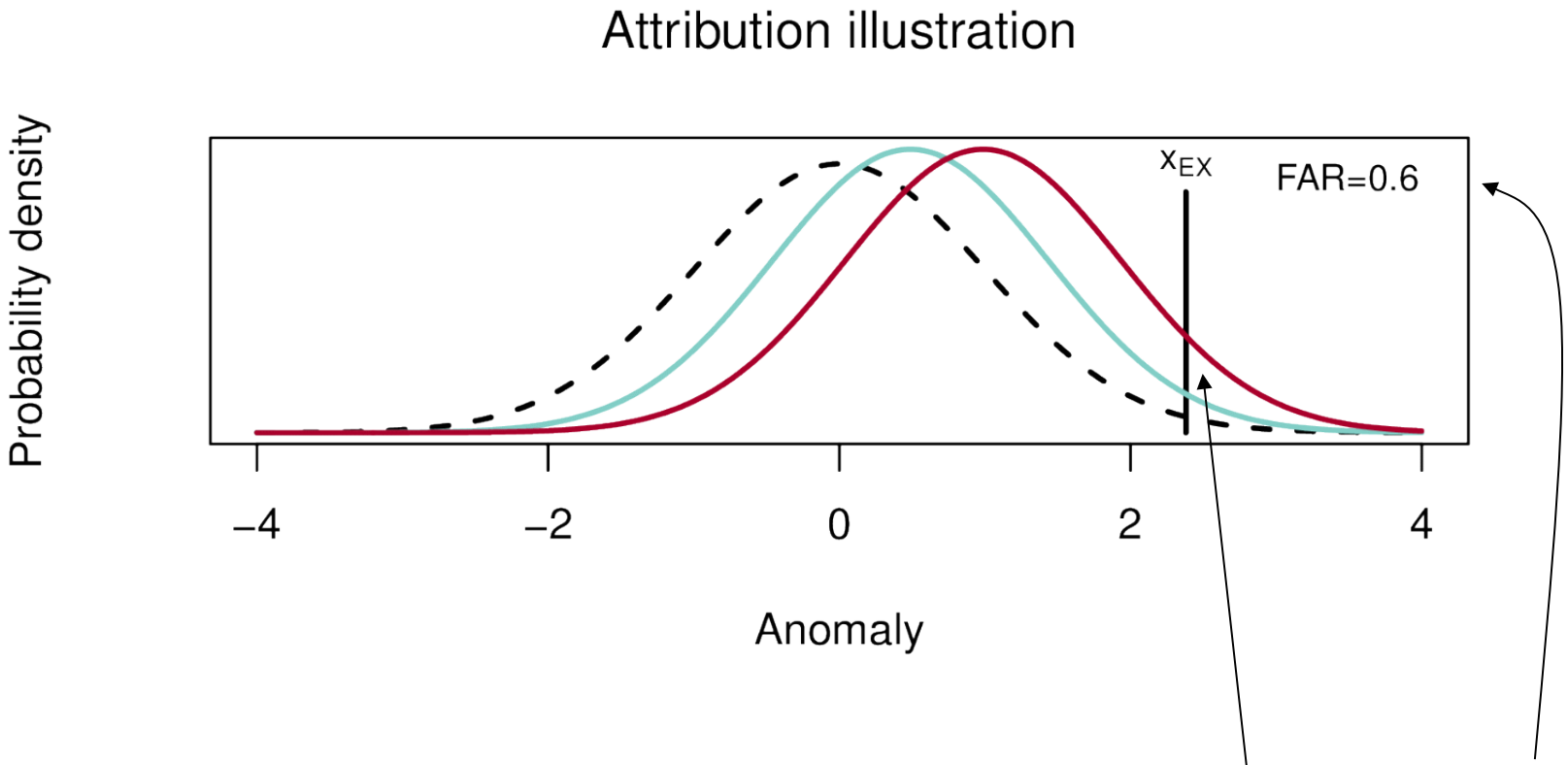


Bellprat and Doblas-Reyes (2016)


alpha – Temperature
90N
70N
50N
30N
10N
10S
30S
50S
180W
140W
100W
60W
20W
0
20E
60E
100E
140E
0
0.25
0.5
0.75
1
1.25
1.5
1.75
2
beta – Temperature
90N
70N
50N
30N
10N
10S
30S
50S
180W
140W
100W
60W
20W
0
20E
60E
100E
140E


Observed Trend JJA 1960 – 2010
Diff Trend Calibrated – Raw
Simulated Trend – Raw JJA 1960 – 2010
Simulated Trend – Calibrated JJA 1960 – 2010
Bias Trend – Raw
Bias Trend – Calibrated

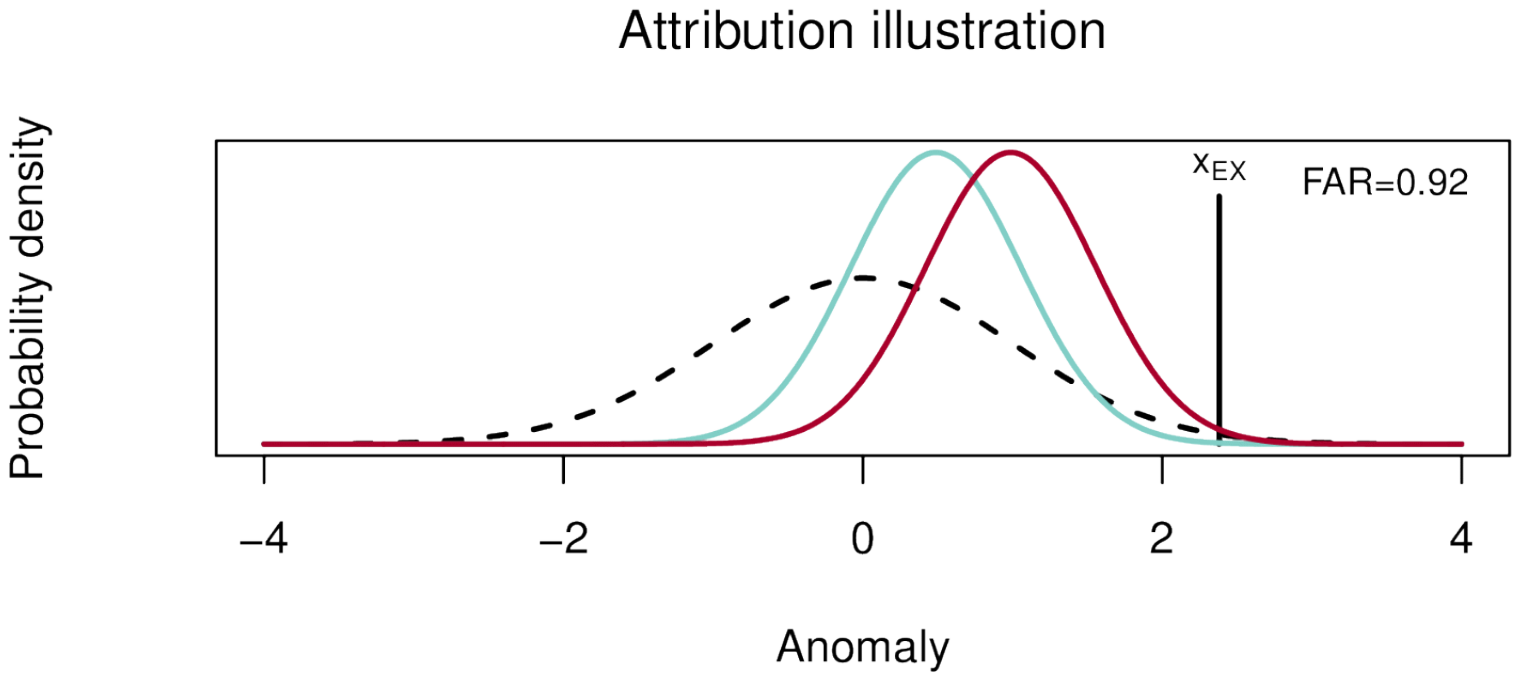
An event attribution case



How ***reliable*** are these probabilities?

Model overconfidence

How can we measure model reliability?

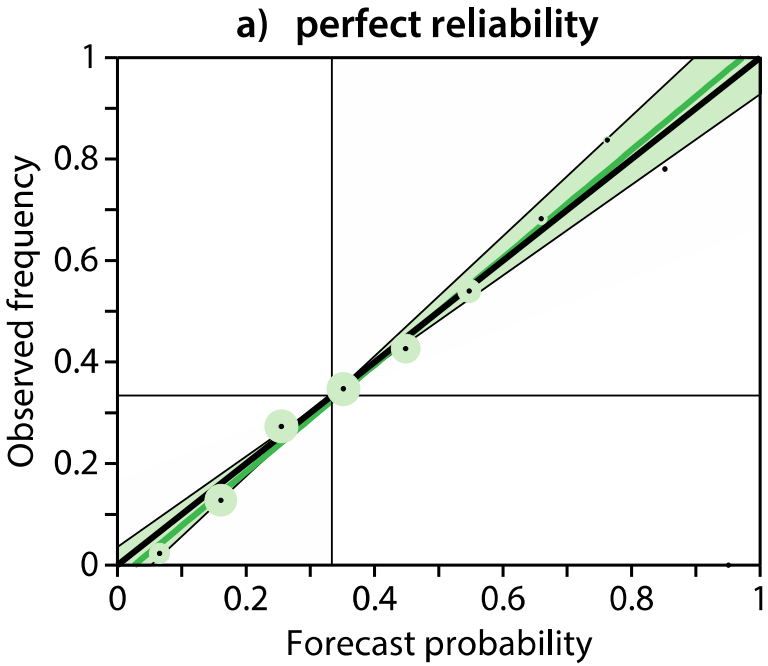


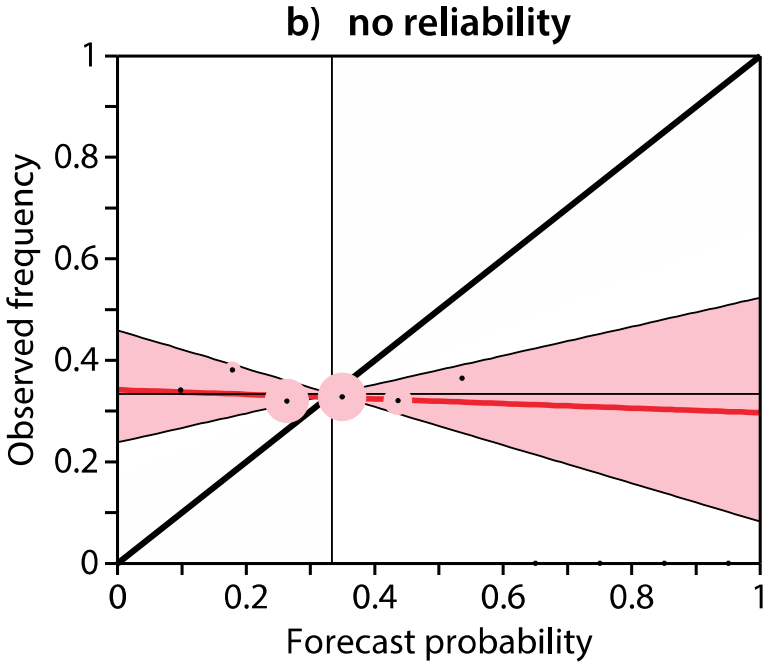
Climate models tend to underestimate climate variability

Bellprat and Doblas-Reyes (2016), GRL

How to measure reliability

When rain > 100 mm is simulated with 80% probability does it actually rain > 100 mm in 80% of the times?





Weisheimer and Palmer (2014), QJRMS

How to correct reliability

Reliability can be corrected by ensemble inflation

Calibrated Hindcast

$$\widehat{F_{t,e}} = \alpha \overline{F_t} + \beta F_{t,e}$$

Variability Observations

$$\alpha = |\rho| \frac{\sigma_o}{\sigma_{em}} \qquad \beta = \sqrt{1-\rho^2} \frac{\sigma_o}{\sigma_e}$$

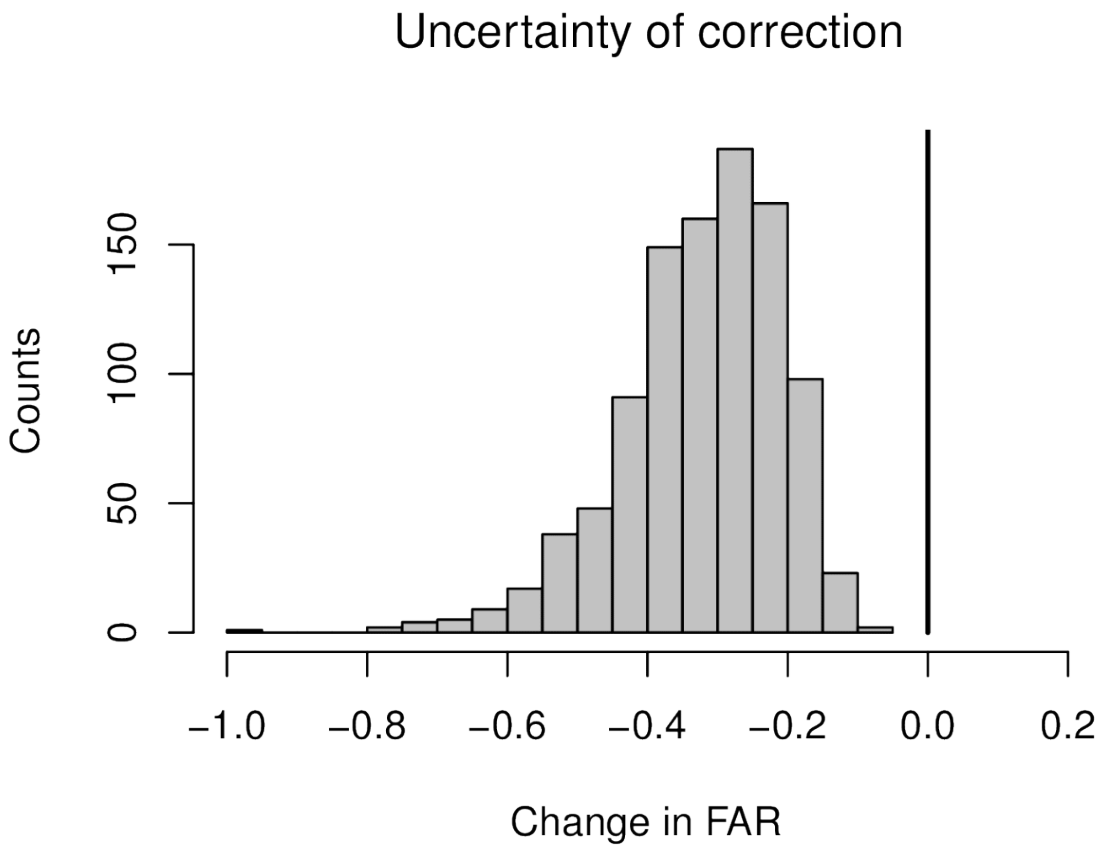
Variability Mean Forecast

Variability Esemble

Von Storch (1999), Doblas-Reyes et al. (2005)

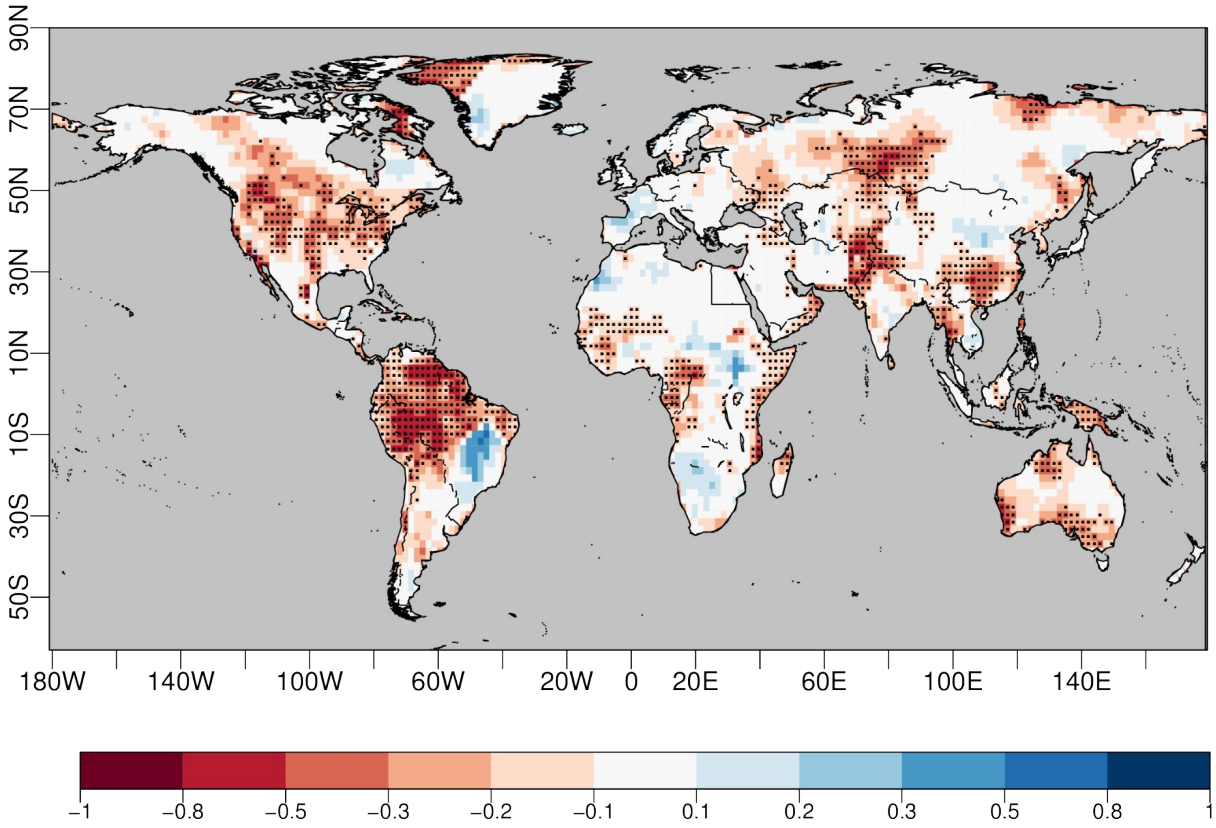
Is the change significant?

Boot-strapping uncertainty of calibration due to limited sample size and uncertain inflation parameters



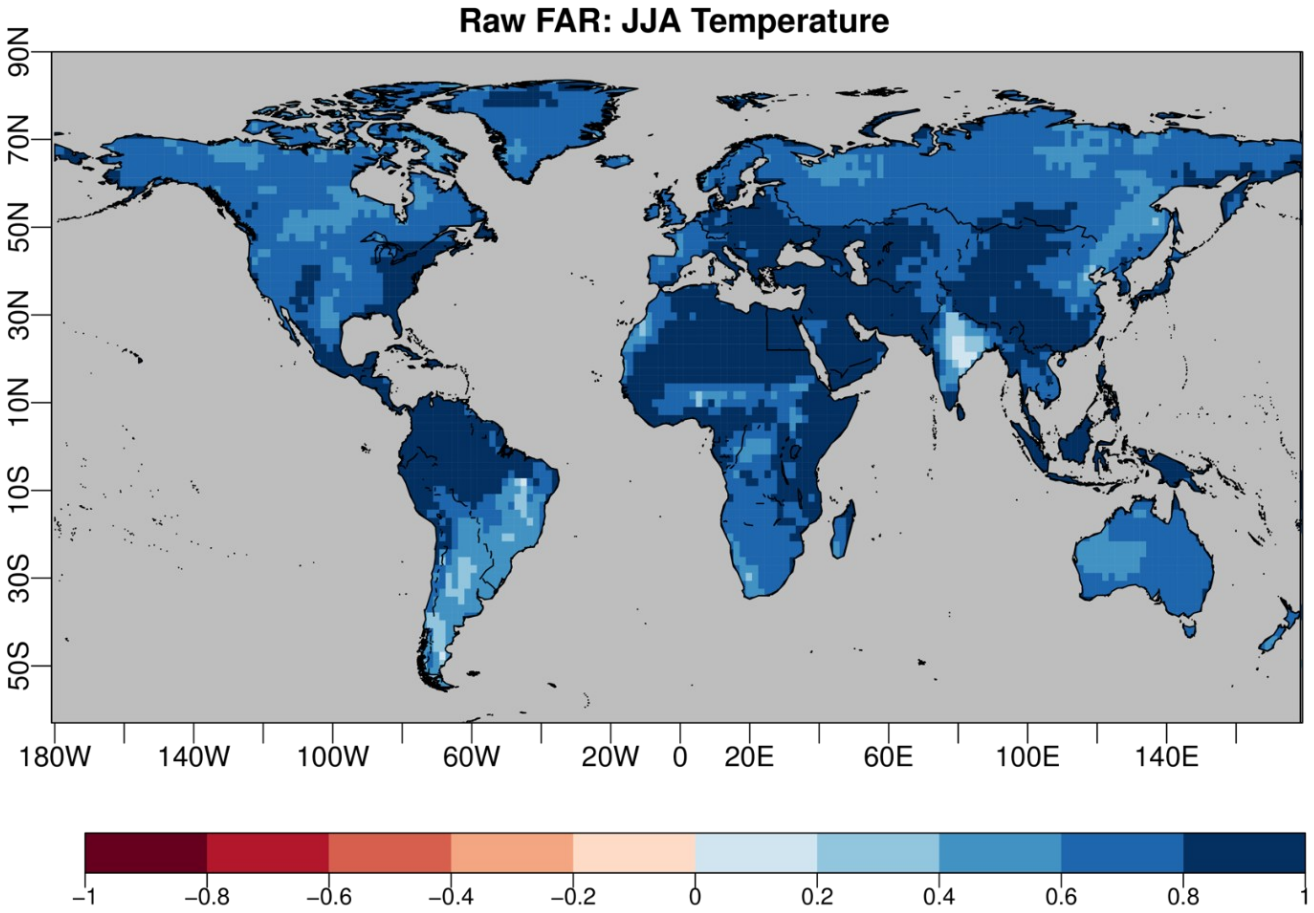
Decrease of FAR globally

Change in attributable risk of hot summers using HadGEM3-A

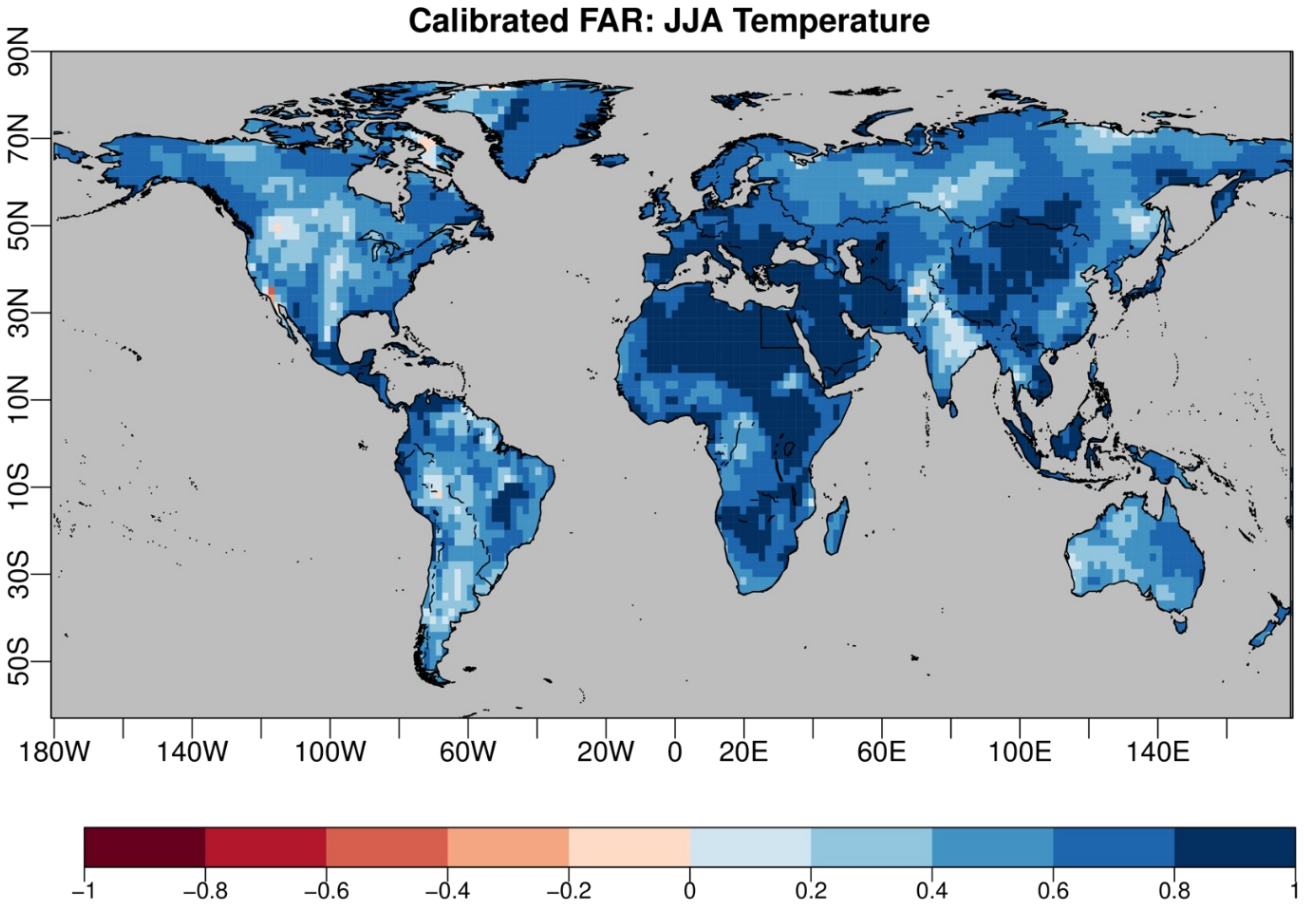


Change in FAR

Raw FAR

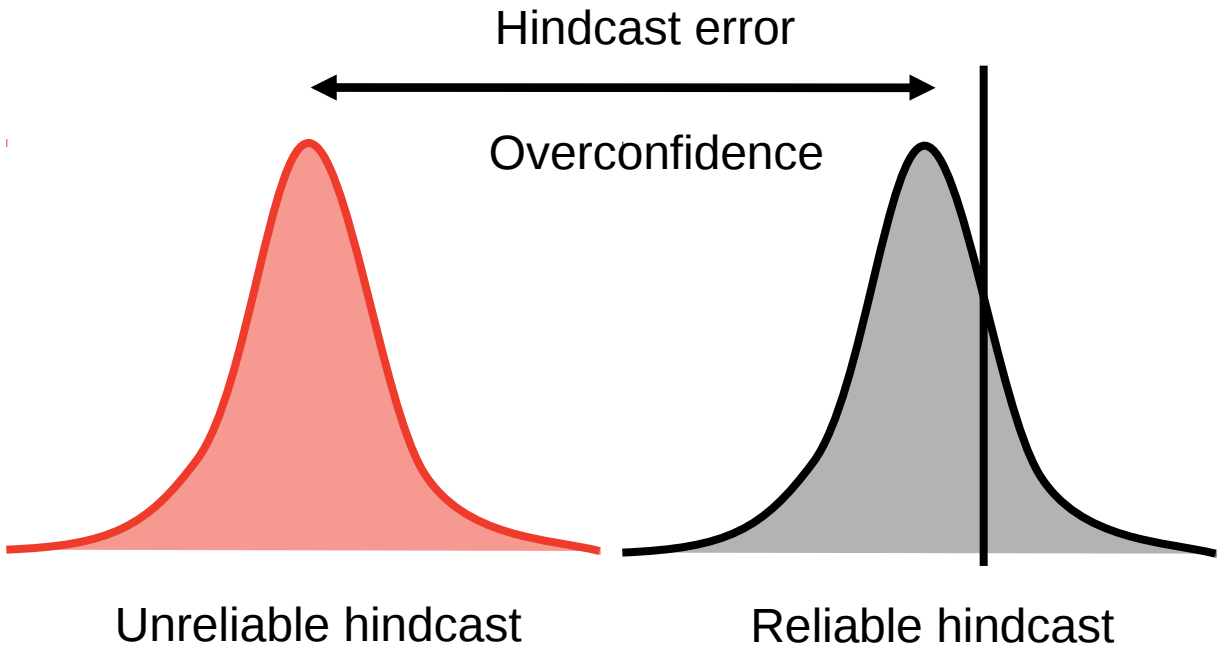

Raw FAR: JJA Temperature
90N
70N
50N
30N
10N
10S
30S
50S
180W
140W
100W
60W
20W
0
20E
60E
100E
140E
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1

Calibrated FAR



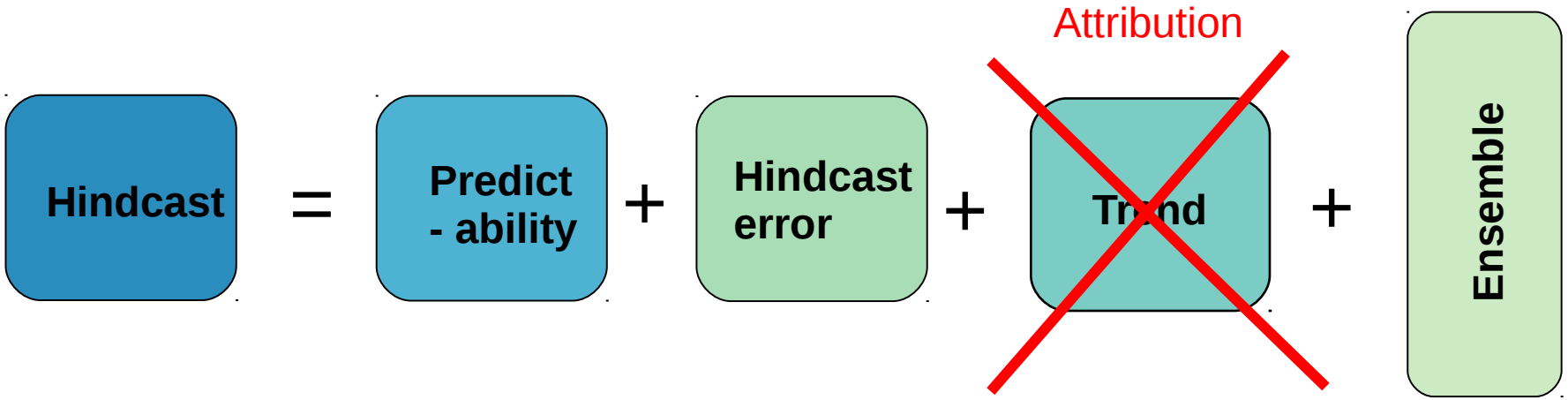
Condition for reliable forecast

Reliability ensured if spread samples model error



Same conditions as RMSE = Spread relation

Toymodel to study reliability and attribution



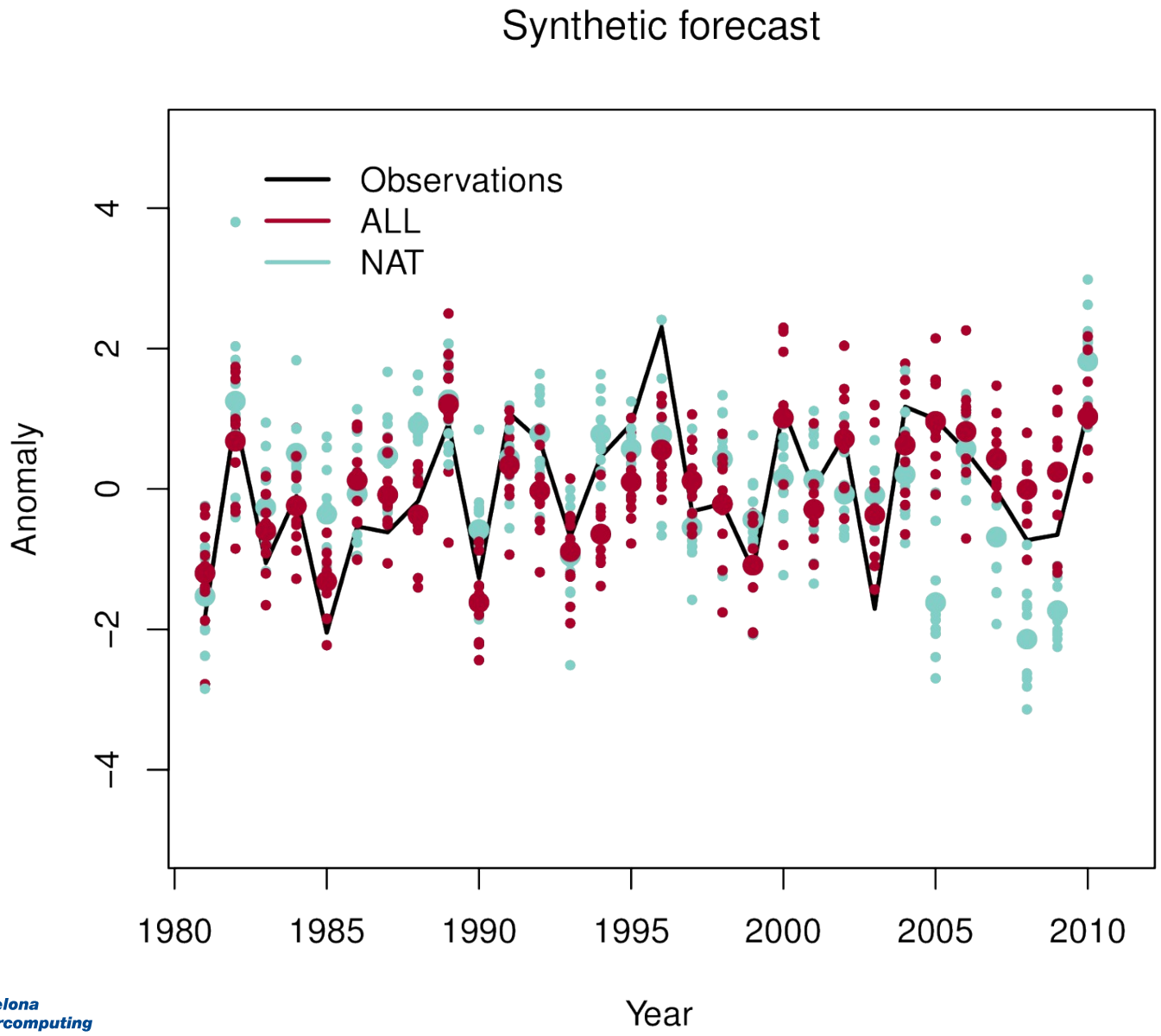
Observations $x_t = x'_t + st$

Model $F_t = \alpha x'_t + \epsilon_\beta + st + (\epsilon_1, \ldots, \epsilon_M)$

Model error $\epsilon_\beta \sim \mathrm{N}(0, \sigma_E = \beta)$

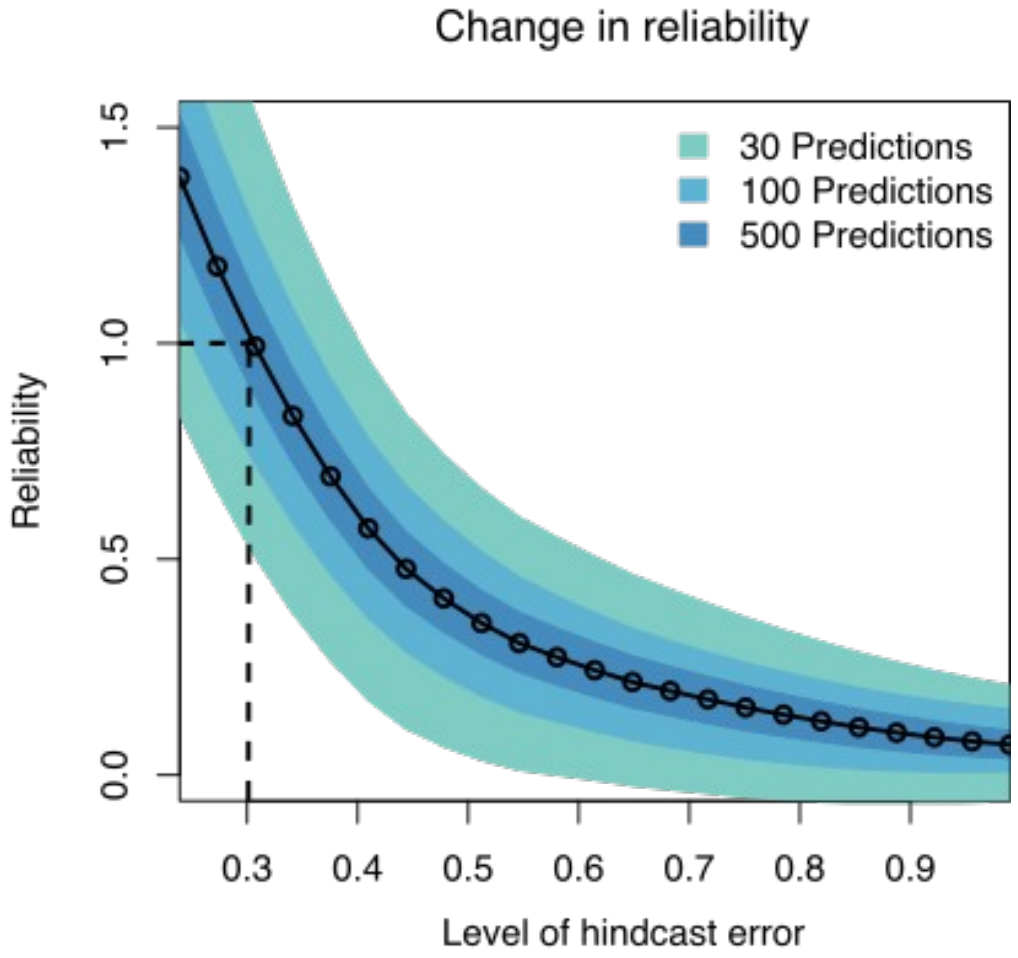
Ensemble spread $\epsilon_M \sim N(0, \sigma_M = \sqrt{(1 - \alpha^2 - \beta^2)})$

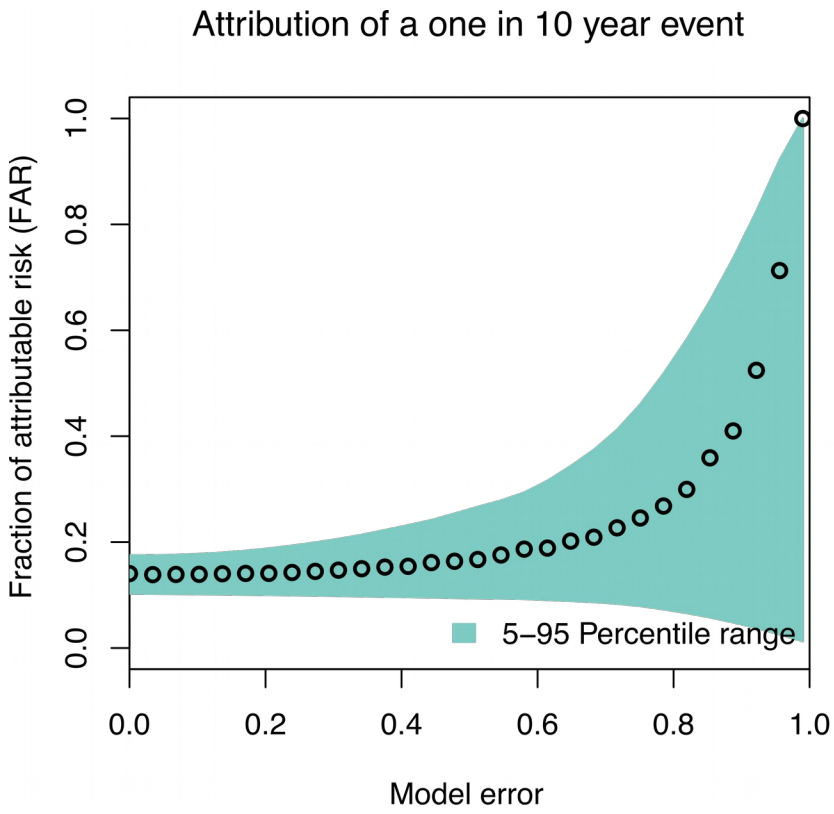
Example toymodel for attribution

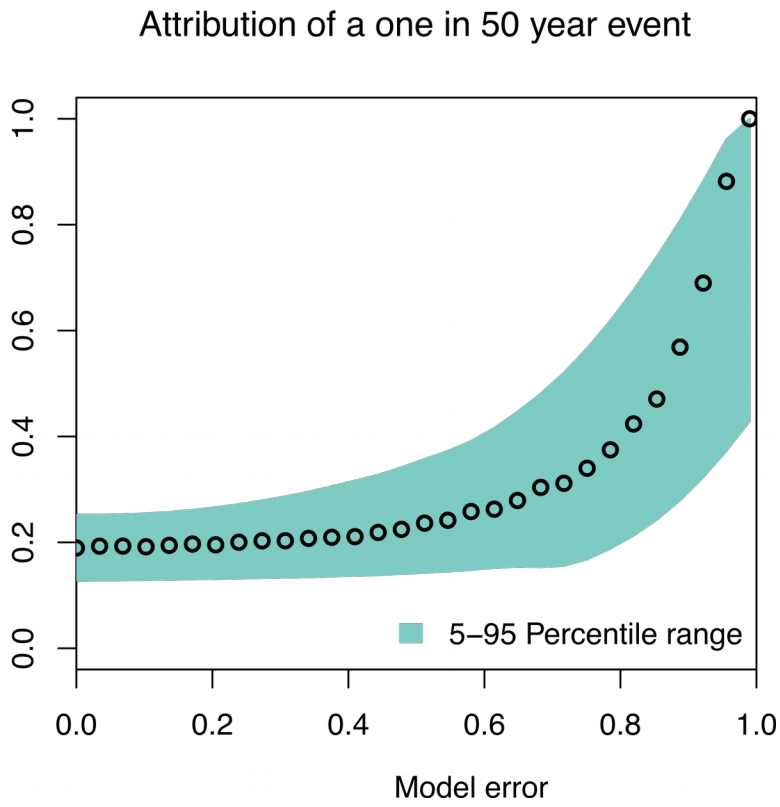


Reliability and sample size

Reliability can be varied at any level, 0=no reliability, 1=perfect




Attribution of a one in 10 year event
Fraction of attributable risk (FAR)
0.0
0.2
0.4
0.6
0.8
1.0
5–95 Percentile range
Model error


Attribution of a one in 50 year event
0.0
0.2
0.4
0.6
0.8
1.0
5–95 Percentile range
Model error

Perfect calibration in toymodel

A reliable model must sample its hindcast error.

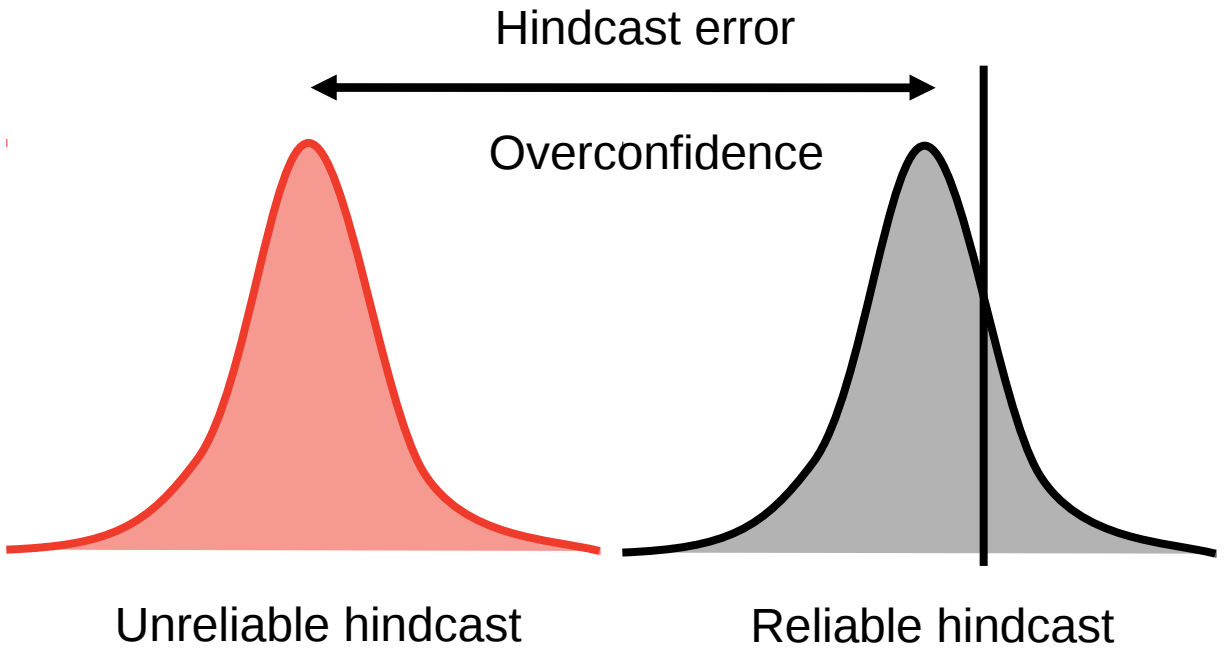
$$\sigma_M^2 = \sigma_E^2 \rightarrow (1-\alpha)^2 + {\beta_r}^2 = 1 - \alpha^2 - {\beta_r}^2$$

$$\beta_r = \sqrt{(\alpha - \alpha^2)} = 0.3 \; (\alpha = 0.1)$$

Can be done in a real model, we can‘t distinguish between β and α. Models need to be calibrated using observations.

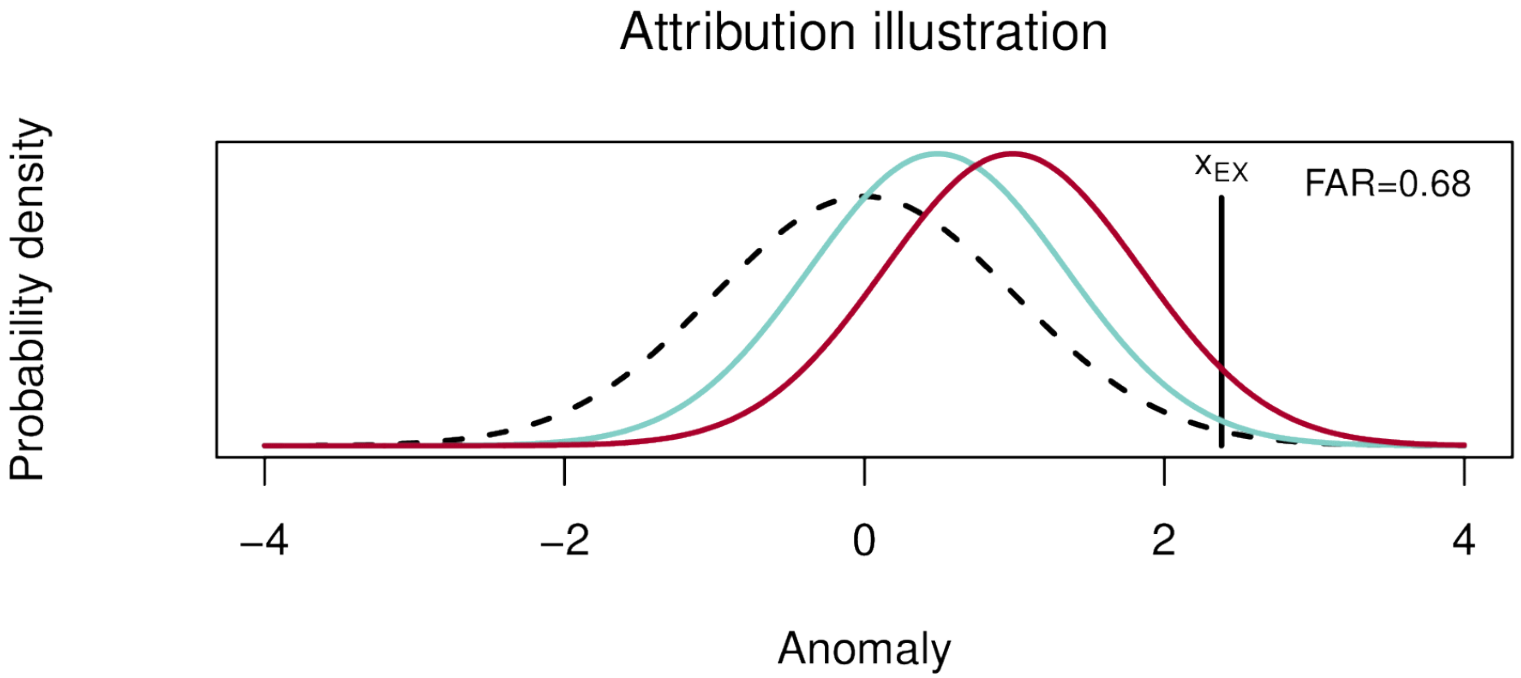
Condition for reliable forecast

Reliability ensured if spread samples model error

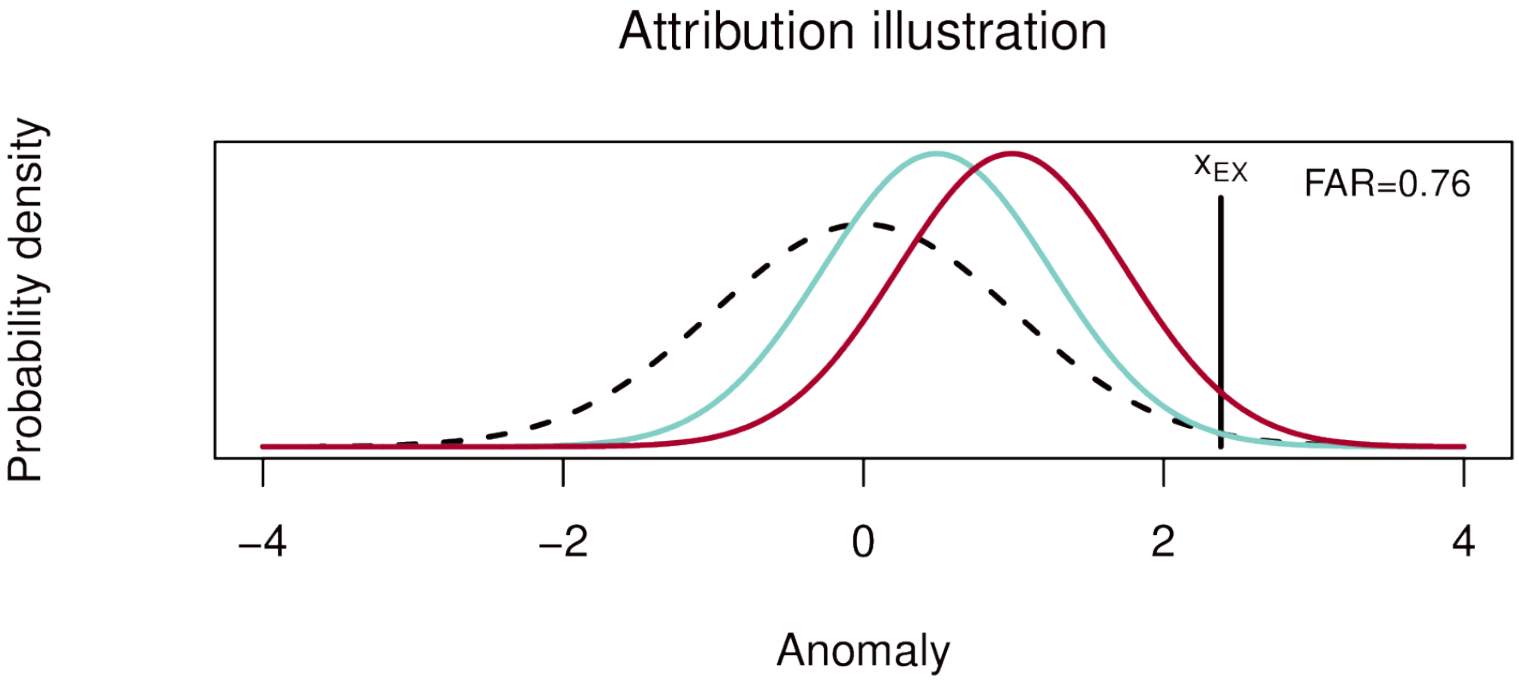


Same conditions as RMSE = Spread relation

Model overconfidence



Model overconfidence



Model overconfidence